

RECEIVED

Page 1 of 7

MAR 07 2002

TECH CENTER 1600/2900

1633



TECH CENTER 1600/2900

FEB 08 2002

RECEIVED

1633

RAW SEQUENCE LISTING

DATE: 01/29/2002

PATENT APPLICATION: US/09/917,376

TIME: 12:18:03

Input Set : A:\401974us.app

Output Set: N:\CRF3\01292002\I917376.raw

P5

3 <110> APPLICANT: DING, SHI-YOU  
4 ADNEY, WILLIAM S.  
5 VINZANT, TODD B.  
6 HIMMEL, MICHAEL E.  
8 <120> TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
9 CELLULOLYTICUS  
11 <130> FILE REFERENCE: 40197.4US01  
13 <140> CURRENT APPLICATION NUMBER: 09/917,376  
14 <141> CURRENT FILING DATE: 2001-07-28.  
16 <160> NUMBER OF SEQ ID NOS: 7  
18 <170> SOFTWARE: PatentIn Ver. 2.1  
20 <210> SEQ ID NO: 1  
21 <211> LENGTH: 957  
22 <212> TYPE: PRT  
23 <213> ORGANISM: Acidothermus cellulolyticus  
25 <220> FEATURE:  
26 <221> NAME/KEY: MOD\_RES  
27 <222> LOCATION: (957)  
28 <223> OTHER INFORMATION: Any amino acid  
30 <400> SEQUENCE: 1  
31 Met Asp Arg Ser Glu Asn Ile Arg Leu Thr Met Arg Ser Arg Arg Leu  
32 1 5 10 15  
34 Val Ser Leu Leu Ala Ala Thr Ala Ser Phe Ala Val Ala Ala Ala Leu  
35 20 25 30  
37 Gly Val Leu Pro Ile Ala Ile Thr Ala Ser Pro Ala His Ala Ala Thr  
38 35 40 45  
40 Thr Gln Pro Tyr Thr Trp Ser Asn Val Ala Ile Gly Gly Gly Gly Phe  
41 50 55 60  
43 Val Asp Gly Ile Val Phe Asn Glu Gly Ala Pro Gly Ile Leu Tyr Val  
44 65 70 75 80  
46 Arg Thr Asp Ile Gly Gly Met Tyr Arg Trp Asp Ala Ala Asn Gly Arg  
47 85 90 95  
49 Trp Ile Pro Leu Leu Asp Trp Val Gly Trp Asn Asn Trp Gly Tyr Asn  
50 100 105 110  
52 Gly Val Val Ser Ile Ala Ala Asp Pro Ile Asn Thr Asn Lys Val Trp  
53 115 120 125  
55 Ala Ala Val Gly Met Tyr Thr Asn Ser Trp Asp Pro Asn Asp Gly Ala  
56 130 135 140  
58 Ile Leu Arg Ser Ser Asp Gln Gly Ala Thr Trp Gln Ile Thr Pro Leu  
59 145 150 155 160  
61 Pro Phe Lys Leu Gly Gly Asn Met Pro Gly Arg Gly Met Gly Glu Arg  
62 165 170 175  
64 Leu Ala Val Asp Pro Asn Asn Asp Asn Ile Leu Tyr Phe Gly Ala Pro

ENTERED

## RAW SEQUENCE LISTING

DATE: 01/29/2002

PATENT APPLICATION: US/09/917,376

TIME: 12:18:03

Input Set : A:\401974us.app

Output Set: N:\CRF3\01292002\I917376.raw

```

65          180          185          190
67 Ser Gly Lys Gly Leu Trp Arg Ser Thr Asp Ser Gly Ala Thr Trp Ser
68          195          200          205
70 Gln Met Thr Asn Phe Pro Asp Val Gly Thr Tyr Ile Ala Asn Pro Thr
71          210          215          220
73 Asp Thr Thr Gly Tyr Gln Ser Asp Ile Gln Gly Val Val Trp Val Ala
74 225          230          235          240
76 Phe Asp Lys Ser Ser Ser Leu Gly Gln Ala Ser Lys Thr Ile Phe
77          245          250          255
79 Val Gly Val Ala Asp Pro Asn Asn Pro Val Phe Trp Ser Arg Asp Gly
80          260          265          270
82 Gly Ala Thr Trp Gln Ala Val Pro Gly Ala Pro Thr Gly Phe Ile Pro
83          275          280          285
85 His Lys Gly Val Phe Asp Pro Val Asn His Val Leu Tyr Ile Ala Thr
86          290          295          300
88 Ser Asn Thr Gly Gly Pro Tyr Asp Gly Ser Ser Gly Asp Val Trp Lys
89 305          310          315          320
91 Phe Ser Val Thr Ser Gly Thr Trp Thr Arg Ile Ser Pro Val Pro Ser
92          325          330          335
94 Thr Asp Thr Ala Asn Asp Tyr Phe Gly Tyr Ser Gly Leu Thr Ile Asp
95          340          345          350
97 Arg Gln His Pro Asn Thr Ile Met Val Ala Thr Gln Ile Ser Trp Trp
98          355          360          365
100 Pro Asp Thr Ile Ile Phe Arg Ser Thr Asp Gly Gly Ala Thr Trp Thr
101          370          375          380
103 Arg Ile Trp Asp Trp Thr Ser Tyr Pro Asn Arg Ser Leu Arg Tyr Val
104 385          390          395          400
106 Leu Asp Ile Ser Ala Glu Pro Trp Leu Thr Phe Gly Val Gln Pro Asn
107          405          410          415
109 Pro Pro Val Pro Ser Pro Lys Leu Gly Trp Met Asp Glu Ala Met Ala
110          420          425          430
112 Ile Asp Pro Phe Asn Ser Asp Arg Met Leu Tyr Gly Thr Gly Ala Thr
113          435          440          445
115 Leu Tyr Ala Thr Asn Asp Leu Thr Lys Trp Asp Ser Gly Gly Gln Ile
116          450          455          460
118 His Ile Ala Pro Met Val Lys Gly Leu Glu Glu Thr Ala Val Asn Asp
119 465          470          475          480
121 Leu Ile Ser Pro Pro Ser Gly Ala Pro Leu Ile Ser Ala Leu Gly Asp
122          485          490          495
124 Leu Gly Gly Phe Thr His Ala Asp Val Thr Ala Val Pro Ser Thr Ile
125          500          505          510
127 Phe Thr Ser Pro Val Phe Thr Thr Gly Thr Ser Val Asp Tyr Ala Glu
128          515          520          525
130 Leu Asn Pro Ser Ile Ile Val Arg Ala Gly Ser Phe Asp Pro Ser Ser
131          530          535          540
133 Gln Pro Asn Asp Arg His Val Ala Phe Ser Thr Asp Gly Gly Lys Asn
134 545          550          555          560
136 Trp Phe Gln Gly Ser Glu Pro Gly Gly Val Thr Thr Gly Gly Thr Val
137          565          570          575

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/917,376

DATE: 01/29/2002

TIME: 12:18:03

Input Set : A:\401974us.app

Output Set: N:\CRF3\01292002\I917376.raw

```

139 Ala Ala Ser Ala Asp Gly Ser Arg Phe Val Trp Ala Pro Gly Asp Pro
140          580          585          590
142 Gly Gln Pro Val Val Tyr Ala Val Gly Phe Gly Asn Ser Trp Ala Ala
143          595          600          605
145 Ser Gln Gly Val Pro Ala Asn Ala Gln Ile Arg Ser Asp Arg Val Asn
146          610          615          620
148 Pro Lys Thr Phe Tyr Ala Leu Ser Asn Gly Thr Phe Tyr Arg Ser Thr
149 625          630          635          640
151 Asp Gly Gly Val Thr Phe Gln Pro Val Ala Ala Gly Leu Pro Ser Ser
152          645          650          655
154 Gly Ala Val Gly Val Met Phe His Ala Val Pro Gly Lys Glu Gly Asp
155          660          665          670
157 Leu Trp Leu Ala Ala Ser Ser Gly Leu Tyr His Ser Thr Asn Gly Gly
158          675          680          685
160 Ser Ser Trp Ser Ala Ile Thr Gly Val Ser Ser Ala Val Asn Val Gly
161          690          695          700
163 Phe Gly Lys Ser Ala Pro Gly Ser Ser Tyr Pro Ala Val Phe Val Val
164 705          710          715          720
166 Gly Thr Ile Gly Gly Val Thr Gly Ala Tyr Arg Ser Asp Asp Cys Gly
167          725          730          735
169 Thr Thr Trp Val Leu Ile Asn Asp Asp Gln His Gln Tyr Gly Asn Trp
170          740          745          750
172 Gly Gln Ala Ile Thr Gly Asp His Ala Asn Leu Arg Arg Val Tyr Ile
173          755          760          765
175 Gly Thr Asn Gly Arg Gly Ile Val Tyr Gly Asp Ile Gly Gly Ala Pro
176          770          775          780
178 Ser Gly Ser Pro Ser Pro Ser Val Ser Pro Ser Ala Ser Pro Ser Leu
179 785          790          795          800
181 Ser Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser Pro
182          805          810          815
184 Ser Ser Ser Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro
185          820          825          830
187 Ser Pro Ser Arg Ser Pro Ser Pro Ser Ala Ser Pro Ser Pro Ser Ser
188          835          840          845
190 Ser Pro Ser Pro Ser Ser Ser Pro Ser Ser Ser Pro Ser Pro Thr Pro
191          850          855          860
193 Ser Ser Ser Pro Val Ser Gly Gly Val Lys Val Gln Tyr Lys Asn Asn
194 865          870          875          880
196 Asp Ser Ala Pro Gly Asp Asn Gln Ile Lys Pro Gly Leu Gln Val Val
197          885          890          895
199 Asn Thr Gly Ser Ser Ser Val Asp Leu Ser Thr Val Thr Val Arg Tyr
200          900          905          910
202 Trp Phe Thr Arg Asp Gly Gly Ser Thr Leu Val Tyr Asn Cys Asp
203          915          920          925
205 Trp Ala Ala Ile Gly Cys Gly Asn Ile Arg Ala Ser Phe Gly Ser Val
206          930          935          940
208 Asn Pro Ala Thr Pro Thr Ala Asp Thr Tyr Leu Gln Xaa
209 945          950          955
212 <210> SEQ ID NO: 2

```

## RAW SEQUENCE LISTING

DATE: 01/29/2002

PATENT APPLICATION: US/09/917,376

TIME: 12:18:03

Input Set : A:\401974us.app

Output Set: N:\CRF3\01292002\I917376.raw

```

213 <211> LENGTH: 2869
214 <212> TYPE: DNA
215 <213> ORGANISM: Acidothermus cellulolyticus
217 <220> FEATURE:
218 <221> NAME/KEY: modified_base
219 <222> LOCATION: (2869)
220 <223> OTHER INFORMATION: a, c, t, g, other or unknown
222 <400> SEQUENCE: 2
223 atggatcggt cggagaacat ccgtctgact atgagatcac gacgattggt atcactgctc 60
224 gccgccactg cgtcggttcg cgtggccgcc gctctgggag ttctgcccac cgcgataacg 120
225 gcttctcctg cgcacgcggc gacgactcag ccgtacacct ggagcaacgt ggcgatcggg 180
226 gccggcggct ttgtcgacgg gatcgctctt aatgaagggt caccgggaat tctgtacgtg 240
227 cggacggaca tcgggggggat gtatcgatgg gatgccgcca acggggcggg gatccctctt 300
228 ctggattggg tgggatggaa caattggggg tacaacggcg tcgtcagcat tgcggcagac 360
229 ccgatcaata ctaacaagg atgggccgcc gtcggaatgt acaccaacag ctgggaccca 420
230 aacgacggag cgattctccg ctctctgat caggcgcaa cgtggcaa atacggccctg 480
231 ccgttcaagc ttggcggcaa catgccggg cgtggaatgg gcgagcggct tgcggtggat 540
232 ccaaacaatg acaacattct gtatttcggc gcccagagc gcaaagggt ctggagaagc 600
233 acagattccg gcgcgacctg gtcccagatg acgaactttc cggacgtagg cactacatt 660
234 gcaaatccca ctgacacgac cggctatcag agcgatattc aaggcgctgt ctgggtcgct 720
235 ttcgacaagt ctctgcatc gctcgggcaa gcgagtaaga ccatttttgt gggcggtggc 780
236 gatcccaata atccggtctt ctggagcaga gacggcgggc cgacgtggca ggcggtgccg 840
237 ggtgcgccga ccggttcat ccgcacaag ggcgtctttg acccggtcaa ccacgtgctc 900
238 tatattgcca ccagcaatac ggggtggtccg tatgacggga gctccggcga cgtctggaaa 960
239 ttctcggtga cctccgggac atggacgcga atcagcccgg taccttcgac ggacacggcc 1020
240 aacgactact ttggttacag cggcctcact atcgaccgcc agcaccgaa cagcataatg 1080
241 gtggcaaccc agatatcgtg gtggccggac accataatct ttcgagcac cgacggcggg 1140
242 gcgacgtgga cgcggatctg ggattggacg agttatcca atcgaaagctt gcgatatgtg 1200
243 cttgacattt cggcggagcc ttggctgacc ttcgcgctac agccgaatcc tcccgtaacc 1260
244 agtccgaagc tcggctggat ggatgaagcg atggcaatcg atccgttcaa ctctgatcgg 1320
245 atgctctacg gaacaggcgc gacgttgtac gcaacaaatg atctcacgaa gtgggactcc 1380
246 gccggccaga ttcatatcgc gccgatggtc aaaggattgg aggagacggc ggtaaacgat 1440
247 ctcatcagcc cgccgtctgg cgcgccgctc atcagcgctc tcggagacct cggcggttc 1500
248 acccacggcg acgttactgc cgtgccatcg acgatcttca cgtcaccggt gttcacgacc 1560
249 ggcaccagcg tcgactatgc ggaattgaat ccgtcgatca tcgttcgcgc tggaaagttc 1620
250 gatccatcga gccaaaccgaa cgacaggcac gtcgcgttct cgacagacgg cggcaagaac 1680
251 tggttccaag gcagcgaacc tggcggggtg acgacgggcg gcaccgtcgc cgcacggcc 1740
252 gacggtctc gtttcgtctg ggctccggc gatcccggtc agcctgtggt gtacgcagtc 1800
253 ggatttggca actcctgggc tgcttcgcaa ggtgttcccg ccaatgccc gatccgctca 1860
254 gaccgggtga atccaaagac tttctatgcc ctatccaatg gaaccttcta tcgaagcacg 1920
255 gacggcggcg tgacattcca accggtcgcg gccggtcttc cgagcagcgg tgccgtcggg 1980
256 gtcatgttcc acgcggtgcc tggaaaagaa ggcgatctgt ggctcgctgc atcgagcggg 2040
257 ctttaccact caaccaatgg cggcagcagt tggctgcaa tcaccggcgt atcctccgcg 2100
258 gtgaacgtgg gatttggtaa gtctgcgccc ggtcgtcat acccagccgt ctttgcgtc 2160
259 ggcacgatcg gagcggttac gggggcgtag cgtccgacg actgtgggac gacctgggta 2220
260 ctgatcaatg atgaccagca ccaatacgg aattggggac aagcaatcac cgtgacacc 2280
261 gcgaatttac ggcgggtgta catagcgacg aacggccgtg gaattgtata cggggacatt 2340
262 ggtgggtgcg cgtccggatc gccgtctccg tcggtgagtc cgtcggttc gccgagcctg 2400
263 agcccagacc cgagcccag cagctcgcca tcgccgtcgc cgtcgccgag ctcgagtcca 2460

```

## RAW SEQUENCE LISTING

DATE: 01/29/2002

PATENT APPLICATION: US/09/917,376

TIME: 12:18:03

Input Set : A:\401974us.app

Output Set: N:\CRF3\01292002\I917376.raw

264 tcctcgtcgc cgtctccgtc gccgtcacca tcgccgagtc cgtctcggtc tccgtcacca 2520  
 265 tcggcgtcgc cgagcccgtc ttcgtcaccg agcccgctctt cgtcaccgtc ttcgtcgcgc 2580  
 266 agcccaacgc cgtcgtcgtc gccggtgtcg ggtgggggtga aggtgcagta taagaataat 2640  
 267 gattcggcgc cgggtgataa tcagatcaag ccgggtttgc aggtgggtgaa taccgggtcg 2700  
 268 tcgtcgggtg atttgtcgac ggtgacgggt cggtactggt tcaccgggga tgggtggctcg 2760  
 269 tcgacactgg tgtacaactg tgactgggcg gcgatcgggt gtgggaatat ccgcgcctcg 2820  
 270 ttcgggtcgg tgaacccggc gacgccgacg gcggacacct acctgcagn 2869  
 273 <210> SEQ ID NO: 3  
 274 <211> LENGTH: 740  
 275 <212> TYPE: PRT  
 276 <213> ORGANISM: Acidothermus cellulolyticus  
 278 <220> FEATURE:  
 279 <223> OTHER INFORMATION: Catalytic domain GH74  
 281 <400> SEQUENCE: 3  
 282 Ala Thr Thr Gln Pro Tyr Thr Trp Ser Asn Val Ala Ile Gly Gly Gly  
 283 1 5 10 15  
 285 Gly Phe Val Asp Gly Ile Val Phe Asn Glu Gly Ala Pro Gly Ile Leu  
 286 20 25 30  
 288 Tyr Val Arg Thr Asp Ile Gly Gly Met Tyr Arg Trp Asp Ala Ala Asn  
 289 35 40 45  
 291 Gly Arg Trp Ile Pro Leu Leu Asp Trp Val Gly Trp Asn Asn Trp Gly  
 292 50 55 60  
 294 Tyr Asn Gly Val Val Ser Ile Ala Ala Asp Pro Ile Asn Thr Asn Lys  
 295 65 70 75 80  
 297 Val Trp Ala Ala Val Gly Met Tyr Thr Asn Ser Trp Asp Pro Asn Asp  
 298 85 90 95  
 300 Gly Ala Ile Leu Arg Ser Ser Asp Gln Gly Ala Thr Trp Gln Ile Thr  
 301 100 105 110  
 303 Pro Leu Pro Phe Lys Leu Gly Gly Asn Met Pro Gly Arg Gly Met Gly  
 304 115 120 125  
 306 Glu Arg Leu Ala Val Asp Pro Asn Asn Asp Asn Ile Leu Tyr Phe Gly  
 307 130 135 140  
 309 Ala Pro Ser Gly Lys Gly Leu Trp Arg Ser Thr Asp Ser Gly Ala Thr  
 310 145 150 155 160  
 312 Trp Ser Gln Met Thr Asn Phe Pro Asp Val Gly Thr Tyr Ile Ala Asn  
 313 165 170 175  
 315 Pro Thr Asp Thr Thr Gly Tyr Gln Ser Asp Ile Gln Gly Val Val Trp  
 316 180 185 190  
 318 Val Ala Phe Asp Lys Ser Ser Ser Ser Leu Gly Gln Ala Ser Lys Thr  
 319 195 200 205  
 321 Ile Phe Val Gly Val Ala Asp Pro Asn Asn Pro Val Phe Trp Ser Arg  
 322 210 215 220  
 324 Asp Gly Gly Ala Thr Trp Gln Ala Val Pro Gly Ala Pro Thr Gly Phe  
 325 225 230 235 240  
 327 Ile Pro His Lys Gly Val Phe Asp Pro Val Asn His Val Leu Tyr Ile  
 328 245 250 255  
 330 Ala Thr Ser Asn Thr Gly Gly Pro Tyr Asp Gly Ser Ser Gly Asp Val  
 331 260 265 270  
 333 Trp Lys Phe Ser Val Thr Ser Gly Thr Trp Thr Arg Ile Ser Pro Val

Use of n and/or Xaa has been detected in the Sequence Listing.  
 Review the Sequence Listing to insure a corresponding  
 explanation is presented in the <220> to <223> fields of  
 Each sequence using n or Xaa.

## VERIFICATION SUMMARY

DATE: 01/29/2002

PATENT APPLICATION: US/09/917,376

TIME: 12:18:04

Input Set : A:\401974us.app

Output Set: N:\CRF3\01292002\I917376.raw

L:208 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1

L:270 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2

L:453 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4